

# Microdata: making metadata matter for machines

**Evergreen International Conference 2013**

# Structured data: making metadata matter for machines

**Evergreen International Conference 2013**

# Our favourite metadata: MARC

```
01231cam a22003258i 4500
001 17601378
005 20130124143505.0
008 130124s2013    pau      b      001 0 eng
010     $a    2012046681
020     $a 9781466636101 (hardcover)
040     $a DLC $b eng $c DLC $e rda
042     $a pcc
050 00 $a TK5105.88815 $b .S42745 2013
082 00 $a 025.042/7 $2 23
245 00 $a Semantic web : $b ontology and knowledge ...
263     $a 1303
264 1  $a Hershey, PA : $b Information Science Reference, $c [2013]
300     $a pages cm
336     $a text $2 rdacontent
337     $a unmediated $2 rdamedia
338     $a volume $2 rdacarrier
504     $a Includes bibliographical references and index.
650 0  $a Semantic Web.
700 1  $a Sheth, A. $q (Amit), $d 1959- $4 edt
```

# Rendered for human consumption

Semantic web : ontology and knowledge base enabled tools,

[Sheth, A. \(Amit\), 1959-](#) (Editor).

## Record details

- **ISBN:** 9781466636101 (hardcover)
- **Physical Description:** pages cm
- **Publisher:** Hershey, PA : Information Science Reference, [2013]

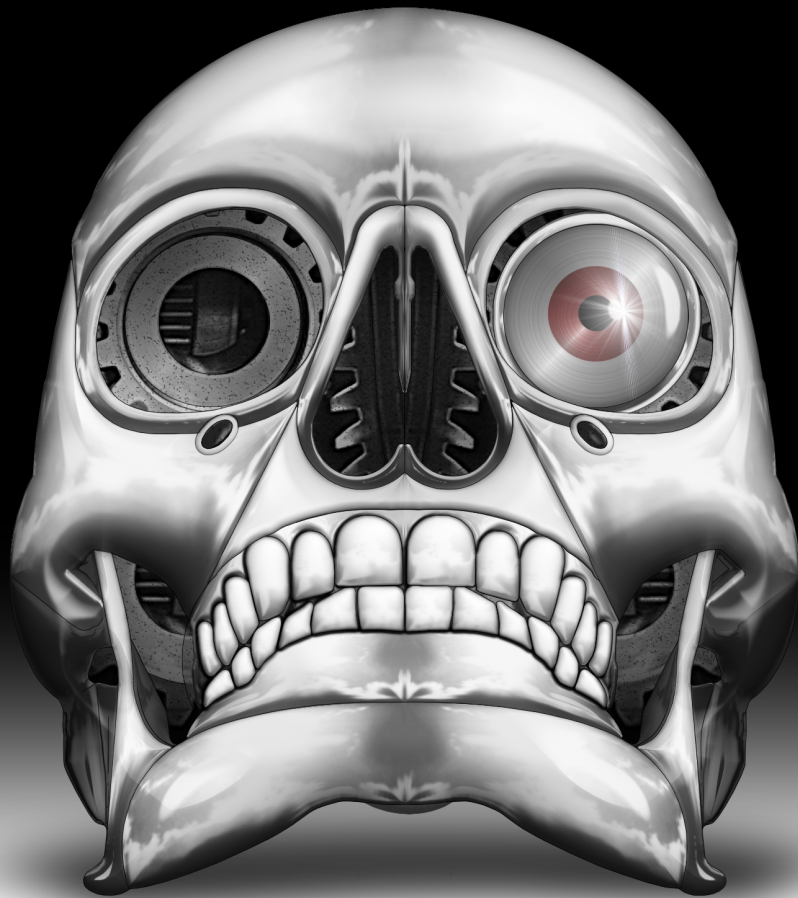
## Content descriptions

**Bibliography, etc. Note:** Includes bibliographical references and index.

## Search for related items by subject

**Subject:** [Semantic Web.](#)

# What about the frakkin' toasters?



DonkeyHotey

*MARC  
is not  
machine  
readable*

Skull by DonkeyHotey



# Linked data basics

- A web page refers to many entities (people, places, events, creative works)
  - Regular HTML links from an entity to another web page are one way of identifying relationships that humans and machines can understand
- A goal of “linked data” is to enable humans and machines to learn more about these entities
  - An *ontology* describes objects, attributes, and their relationships (MusicAlbum → byArtist → MusicGroup)
- “Knowledge Graph” in Google search results is one concrete example of linked data

# Enter schema.org microdata

- Search engines wanted better metadata than plain HTML could provide
- Semantic web was not evolving in practice
- June 2, 2011: Bing, Google, and Yahoo **announce** schema.org:

*“to create and support a standard set of schemas for structured data markup on web pages”*

- April 6, 2013: Dan Brickley announces 1.0a revision

# Human view

Semantic web : ontology and knowledge base enabled tools,

[Sheth, A. \(Amit\), 1959-](#) (Editor).

## Record details

- **ISBN:** 9781466636101 (hardcover)
- **Physical Description:** pages cm
- **Publisher:** Hershey, PA : Information Science Reference, [2013]

## Content descriptions

**Bibliography, etc. Note:** Includes bibliographical references and index.

## Search for related items by subject

**Subject:** [Semantic Web.](#)



# Plain HTML

```
<div>
  <abbr class="unapi-id" title="tag:localhost,2013:biblio-
record_entry/211"></abbr>

  <div><div>
    <h1>
      Semantic web : ontology and knowledge base enabled tools,
      services, and applications / Amit Sheth, editor.</h1>
    <div>
      <span>
        <a href="/eg/opac/results?query=%20Sheth%20%A">
          <span>Sheth, A. (Amit), 1959-</span>
          </a> (Editor).
        </span>
      </div>
    </div>
  </div>
</div>
```

# HTML + schema.org (microdata)

```
<div itemscope itemtype="http://schema.org/Book">
  <abbr class="unapi-id" title="tag:localhost,2013:biblio-
record_entry/211"></abbr>

  <div><div>
    <h1 itemprop="name">
      Semantic web : ontology and knowledge base enabled tools,
      services, and applications / Amit Sheth, editor.</h1>
    <div>
      <span>
        <a href="/eg/opac/results?query=%20Sheth%20A"
          itemtype="http://schema.org/Person"
          itemprop="contributor">
          <span itemprop="name">Sheth, A. (Amit),</span>
          <span itemprop="birthdate">1959</span>
          - </a> (Editor).
        </span>
      </div></div></div></div>
```

# Search engine view

## Extracted structured data

### Item

<b>type:</b>	<a href="http://schema.org/book">http://schema.org/book</a>
<b>property:</b>	
name:	Semantic web : ontology and knowledge base enabled tools, services, and applications / Amit Sheth, editor.
contributor:	<i>Item 1</i>
isbn:	9781466636101 (hardcover)
keywords:	Includes bibliographical references and index.
keywords:	Semantic Web.

### Item 1

<b>type:</b>	<a href="http://schema.org/person">http://schema.org/person</a>
<b>property:</b>	
name:	Sheth, A. (Amit),
birthdate:	1959

# So. Wat.

- Things, not strings:
  - For differentiation:
    - “Dan Wells” the author vs. the famous Evergreen developer
  - For similarity:
    - “West Side Story” as an adaptation of “Romeo & Juliet”
- Exposes and enhances library resources:
  - Search engine results could link to local or preferred libraries
  - Physical *or* electronic resources
    - Ross Singer's [Backbeat](#) GreaseMonkey script

# BackBeat



Language: English (US)

Change

Your Account Log in

[Library Home](#) [Electronic Journals \(A-Z\)](#) [Databases](#) [Archives](#) [Course Reserves](#) [Feedback](#)

Search the Catalog [Advanced Search](#)

Search for

Keyword

in

[◀ Search Results](#) Showing Item 6 of 1007 [◀ Previous](#) [Next ▶](#)

## Trumpet concertos

[Telemann, Georg Philipp, 1681-1767](#) (Author). [Boughton, William](#) (Added Author). [Wallace, English String Orchestra](#) (Added Author).

Backbeat!

Record

- **Physical Description:** 1 sound disc : digital, stereo
- **Publisher:** Ocean, NJ : Musical Heritage Society, 1991.

## Content descriptions

**General Note:** John Wallace, trumpet ; English String Orchestra ; William Boughton, conductor.

**Formatted Contents Note:** Overture in D -- Concerto in D -- Sinfonia in F -- Concerto in D -- Concerto in D -- Concerto in D -- Suite in D -- Concerto in E flat.

## Search for related items by subject



[Add to my list](#)

[Print / Email](#)

[Export to RefWorks](#)

# Telemann, Georg Philipp, :: Trumpet concertos

rdio

spotify

## Georg Philipp Telemann

**album** [Telemann, G.P.: 6 Concertos](#)

**album** [Complete Trumpet Concertos Part: 3](#)

**album** [Complete Trumpet Concertos Part: 1](#)

**album** [Complete Trumpet Concertos Part: 2](#)



# schema.org: beyond Book

- Object types (sampling):
  - Book, Map, Movie, MusicAlbum, Painting, Photograph, Sculpture
- Attributes (sampling):
  - Main/added entries: author, accountablePerson, contributor, creator, editor
  - Publication: copyrightHolder, copyrightYear, dateCreated, datePublished, publisher
  - Description: description, about, audience, award, contentRating, genre, keywords
  - Social: comment, interactionCount, review

# Libraries are objects too

- schema.org vocabulary defines “**Thing > Organization > LocalBusiness > Library**”
- Supports attributes such as:
  - Address and contact info (phone, fax, email)
  - Opening hours
  - Branch relationships
  - Payment types accepted
  - Events
- Search engines could do a lot with this!



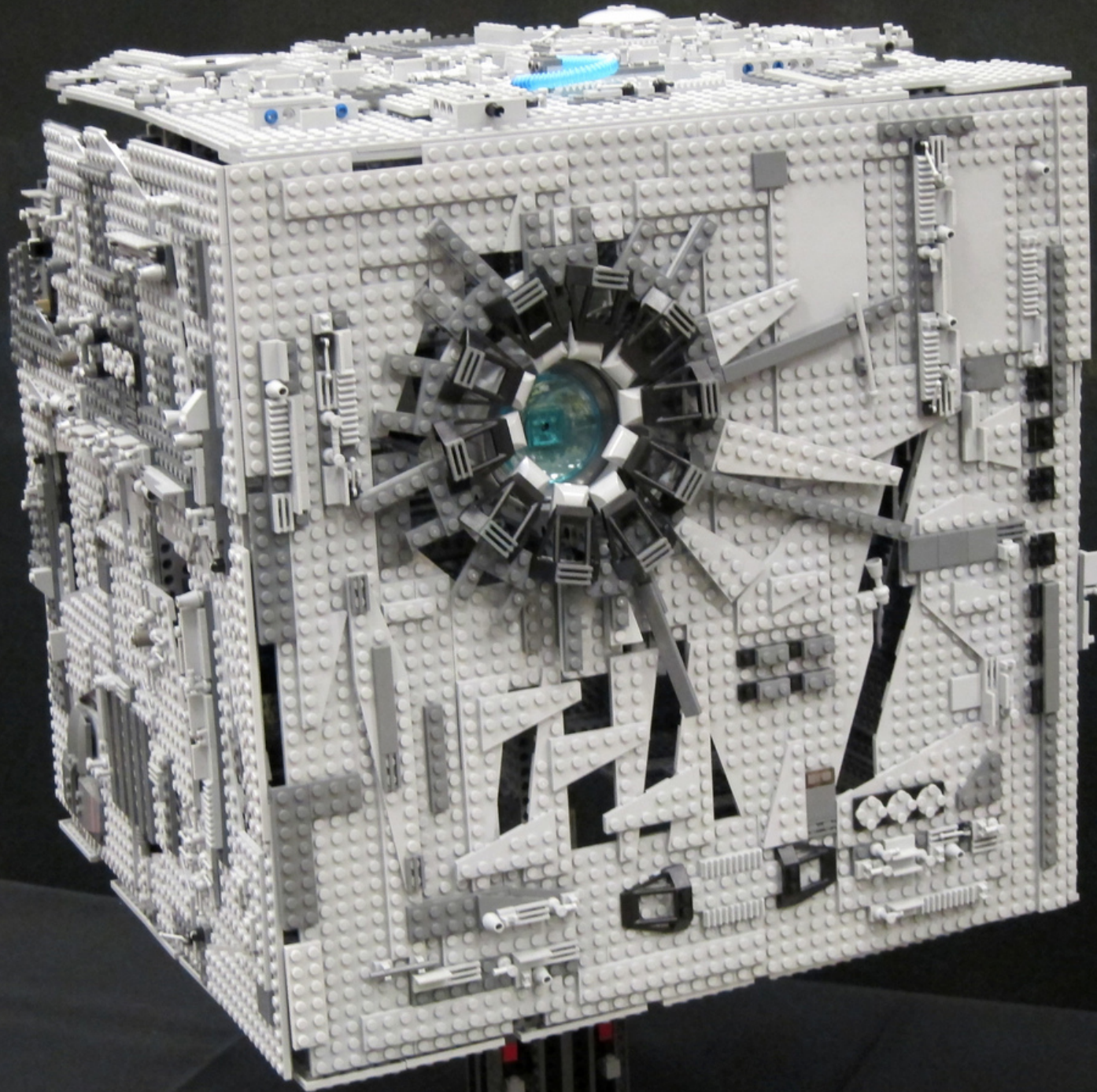
# Linking open data

- Authorized LoC headings could link to <http://id.loc.gov/authorities/...>
  - VIAF: interesting target, but does not link out
- What is the canonical link for a book?
  - **isbn** is a property in schema.org, but:
    - The “same” book can have multiple ISBNs
    - Not everything has an ISBN
  - Just link to Freebase? OpenLibrary? WorldCat?
  - FRBR? WEMI?
  - ARGH?!?

# schema.org: vocab in progress

- Extension proposals: coordinated by the W3C [Semantic Web Interest Group](#) Web Schemas task force
- Bibliographic extension proposals: [W3C Schema Bib Extend Community Group](#)
  - Initiated by Richard Wallis (OCLC) in September 2012 with monthly calls
  - First action agreed upon March 2013: “to promote citation property to CreativeWork”
  - Much discussion about identifiers, FRBR, BibFrame, and holdings
  - Participants include Richard Wallis, Karen Coyle, Jeff Young, Ross Singer, Jodi Schneider, Laura Dawson...

schema.borg?



**Assimilated vocabs:**

- rNews
- LRMI
- GoodRelations

# A sad tale

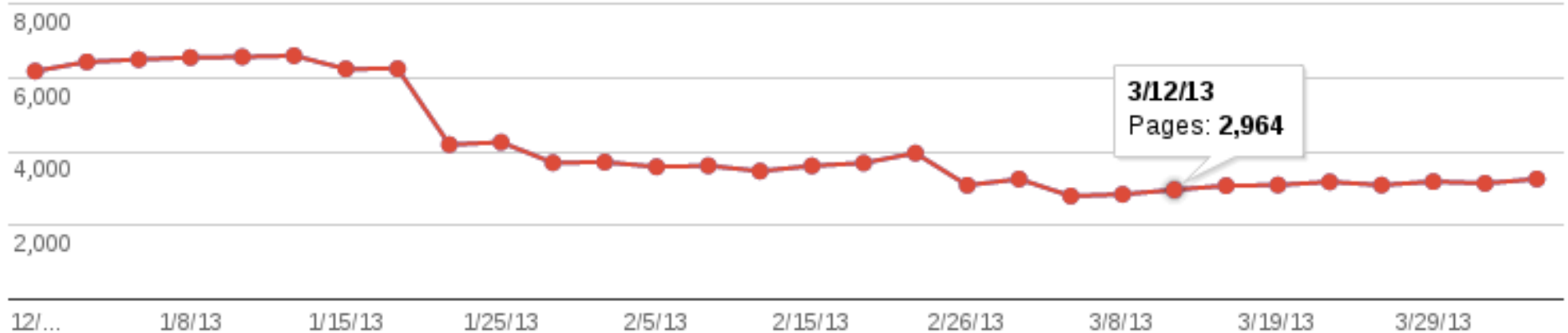
Structured Data > Book (Markup: schema.org)



## Book

■ **3,257** Items

■ **3,257** Pages



# Evergreen: schema.org state

- Evergreen 2.2 through 2.4:
  - Two primary types: Book and MusicRecording\*
  - Just plain text for attributes
- A **working branch** improves this greatly:
  - Primary types: Book, Map, MusicAlbum
  - More granular Organization and Person types for main/added entries
  - Birth and death dates for Person objects



# Evergreen challenges

- Linking out:
  - Works / expressions: where and how?
  - Controlled headings -> linked authorities
- Type mapping from MARC is tough:
  - Is a 2D projected medium a movie, TV show, slide show, or a magic lantern?
- Logic is implemented in TPAC templates
  - That's where we publish HTML, so it makes some sense
  - Also complicates the templates and further hard-codes mappings
- Is it worth the effort?

# RDFa Lite

- Like Microdata, but a W3C standard that is backwards compatible with RDFa
- Currently only lacks equivalent of **itemref** “copy and paste” attribute
- Supported by schema.org founders, *and* Facebook
- A [working branch](#) offers TPAC templates that are 1:1 with current schema.org microdata

# HTML + schema.org (RDFa Lite)

```
<div vocab="http://schema.org" typeof="http://schema.org/Book">
  <abbr class="unapi-id" title="tag:localhost,2013:biblio-
record_entry/211" />

  <div><div>
    <h1 property="name">
      Semantic web : ontology and knowledge base enabled tools,
      services, and applications / Amit Sheth, editor.</h1>
    <div>
      <span>
        <a href="/eg/opac/results?query=%20Sheth%20A"
          typeof="http://schema.org/Person"
          property="contributor">
          <span property="name">Sheth, A. (Amit),</span>
          <span property="birthdate">1959</span>
          - </a> (Editor).
        </span>
      </div></div></div></div>
```



# References

- Articles:

- Berners-Lee, T., Hendler, J., and Lassila, O. (2001) "The Semantic Web". Scientific American, [May](#).
- Coyle, Karen. "Rich Snippets". Coyle's InFormation. <http://kcoyle.blogspot.ca/2012/09/rich-snippets.html>
- Sporny, Manu. "Mythical Differences: RDFa Lite vs. Microdata." The Beautiful, Tormented Machine. <http://manu.sporny.org/2012/mythical-differences/>

- Sites:

- schema.org ontology: <http://schema.org>
- RDFa Lite standard: <http://www.w3.org/TR/rdfa-lite/>
- W3C schema.org bibliographic extension community: <http://www.w3.org/community/schemabibex>
- W3C semantic web community: <http://www.w3.org/standards/semanticweb>

- Tools:

- BackBeat: <https://github.com/rsinger/BackBeat>

# Microdata: making metadata matter for machines

**Evergreen International Conference 2013**

Dan Scott, <http://coffeecode.net>

April 12, 2013



# Structured data: making metadata matter for machines

Evergreen International Conference 2013

Dan Scott, <http://coffeecode.net>

April 12, 2013



To avoid taking a microdata vs. RDFa Lite stance, I'm going to group both of them as “structured data”.

## Our favourite metadata: MARC

```
01231cam a22003258i 4500
001 17601378
005 20130124143505.0
008 130124s2013    pau      b    001 0 eng
010    $a 2012046681
020    $a 9781466636101 (hardcover)
040    $a DLC $b eng $c DLC $e rda
042    $a pcc
050 00 $a TK5105.88815 $b .S42745 2013
082 00 $a 025.042/7 $2 23
245 00 $a Semantic web : $b ontology and knowledge ...
263    $a 1303
264 1  $a Hershey, PA : $b Information Science Reference, $c [2013]
300    $a pages cm
336    $a text $2 rdacontent
337    $a unmediated $2 rdamedia
338    $a volume $2 rdacarrier
504    $a Includes bibliographical references and index.
650 0  $a Semantic Web.
700 1  $a Sheth, A. $q (Amit), $d 1959- $4 edt
```

Now available in extra-tasty RDA flavour!

# Rendered for human consumption

Semantic web : ontology and knowledge base enabled tools,  
[Sheth, A. \(Amit\), 1959-](#) (Editor).

## Record details

- **ISBN:** 9781466636101 (hardcover)
- **Physical Description:** pages cm
- **Publisher:** Hershey, PA : Information Science Reference, [2013]

## Content descriptions

**Bibliography, etc. Note:** Includes bibliographical references and index.

## Search for related items by subject

**Subject:** [Semantic Web.](#)

With some difficulty, we can turn MARC into a reasonable HTML representation that we can make sense of in a Web browser.

# What about the frakkin' toasters?

*MARC  
is not  
machine  
readable*

Skull by DonkeyHotey



Previous efforts include:

Microformats:

unAPI – offers links from HTML rendering to alternative formats (MARCXML, MODS, RIS, ...), but was only used by Evergreen, Zotero, rebase, and Bebo

ContextObjects in Spans, commonly abbreviated COinS, surface OpenURL ContextObjects in a SPAN tag – widely used, primarily for citations

RDFa: proposed in 2004, like Microformats on steroids, allows for implementation of RDF in (X)HTML; did not get traction

# Linked data basics

- A web page refers to many entities (people, places, events, creative works)
  - Regular HTML links from an entity to another web page are one way of identifying relationships that humans and machines can understand
- A goal of “linked data” is to enable humans and machines to learn more about these entities
  - An *ontology* describes objects, attributes, and their relationships (MusicAlbum → byArtist → MusicGroup)
- “Knowledge Graph” in Google search results is one concrete example of linked data

## Enter schema.org microdata

- Search engines wanted better metadata than plain HTML could provide
- Semantic web was not evolving in practice
- June 2, 2011: Bing, Google, and Yahoo **announce** schema.org:  
*“to create and support a standard set of schemas for structured data markup on web pages”*
- April 6, 2013: Dan Brickley announces 1.0a revision

Like Arthur C Clarke's “billions and billions” quote, semantic web activity seemed to focus on celebrating quantities of RDF triples but most people generating web pages didn't see any value; RDFa was struggling to gain traction

Counterpoint: Manu Sporny argued that RDFa was emerging in mid-2009:  
<http://blog.whatwg.org/microdata1#comment-40780>)

Once schema.org launched, it gained adopters and the vocabulary continued to evolve.

In 2012, the schema.org process opened up to allow extension proposals from outside the founding partners



# Human view

Semantic web : ontology and knowledge base enabled tools,  
[Sheth, A. \(Amit\), 1959-](#) (Editor).

## Record details

- **ISBN:** 9781466636101 (hardcover)
- **Physical Description:** pages cm
- **Publisher:** Hershey, PA : Information Science Reference, [2013]

## Content descriptions

**Bibliography, etc. Note:** Includes bibliographical references and index.

## Search for related items by subject

**Subject:** [Semantic Web.](#)

With some difficulty, we can turn MARC into a reasonable HTML representation that we can make sense of in a Web browser.

# Plain HTML

```
<div>
  <abbr class="unapi-id" title="tag:localhost,2013:biblio-
record_entry/211"></abbr>

  <div><div>
    <h1>
      Semantic web : ontology and knowledge base enabled tools,
      services, and applications / Amit Sheth, editor.</h1>
    <div>
      <span>
        <a href="/eg/opac/results?query=%20Sheth%20%A">
          <span>Sheth, A. (Amit), 1959-</span>
        </a> (Editor).
      </span>
    </div>
  </div>
</div>
</div>
```

An HTML page from today's TPAC, stripped of id and class attributes, and also stripped of the existing schema.org microdata that we have produced since Evergreen 2.2.

# HTML + schema.org (microdata)

```
<div itemscope itemtype="http://schema.org/Book">
  <abbr class="unapi-id" title="tag:localhost,2013:biblio-
record_entry/211"></abbr>

  <div><div>
    <h1 itemprop="name">
      Semantic web : ontology and knowledge base enabled tools,
      services, and applications / Amit Sheth, editor.</h1>
    <div>
      <span>
        <a href="/eg/opac/results?query=%20Sheth%20A"
          itemprop="http://schema.org/Person"
          itemprop="contributor">
          <span itemprop="name">Sheth, A. (Amit),</span>
          <span itemprop="birthdate">1959</span>
          - </a> (Editor).
        </span>
      </div></div></div></div>
```

Schema.org microdata annotates the existing HTML with additional attributes that assert that a given HTML element is an object, or a property of a containing object..

In some cases, adding more granular HTML elements rewards us with the ability to provide more specific properties (in this case, separating out the author's birth date from their name).

# Search engine view

## Extracted structured data

### Item

<b>type:</b>	http://schema.org/book
<b>property:</b>	
name:	Semantic web : ontology and knowledge base enabled tools, services, and applications / Amit Sheth, editor.
contributor:	Item 1
isbn:	9781466636101 (hardcover)
keywords:	Includes bibliographical references and index.
keywords:	Semantic Web.

### Item 1

<b>type:</b>	http://schema.org/person
<b>property:</b>	
name:	Sheth, A. (Amit),
birthdate:	1959

*Courtesy of the [Google Rich Snippets Tool](#)*

Given these assertions, a search engine crawling the page can be highly confident that the page is primarily about this particular book, as well as a relationship between the author of the book and the book itself.

## So. Wat.

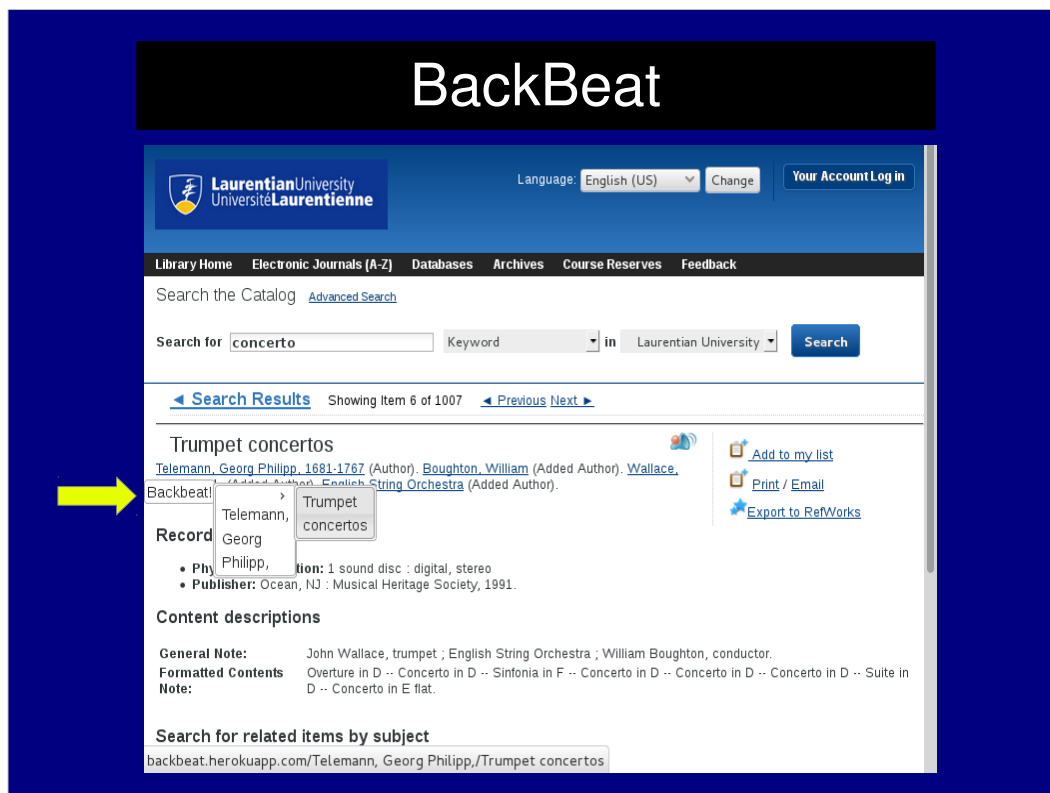
- Things, not strings:
  - For differentiation:
    - “Dan Wells” the author vs. the famous Evergreen developer
  - For similarity:
    - “West Side Story” as an adaptation of “Romeo & Juliet”
- Exposes and enhances library resources:
  - Search engine results could link to local or preferred libraries
  - Physical *or* electronic resources
    - Ross Singer's [Backbeat](#) GreaseMonkey script

Currently Google Books search results include an option to look for resources at libraries via WorldCat... but many libraries have not exposed their holdings in WorldCat.

Note that “preferences” are mostly hypothetical; beyond Google Scholar, there doesn't seem to be much search engine interest in this. Yet.

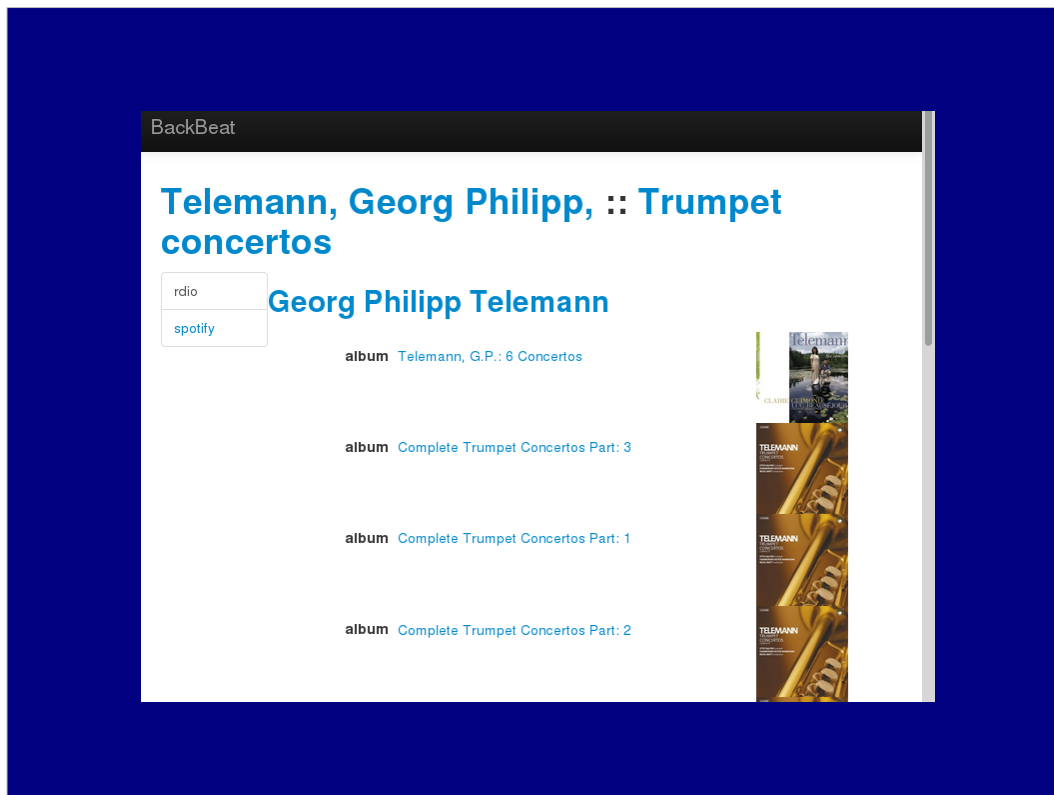
Well... a few years ago, some Googlers were interested in crawling Evergreen's catalogue. But the JSPAC was unfriendly for that purpose, and the basic catalogue was... not good. **Note to self:** strike up that conversation again.

Maybe we can convince Yahoo to give it a try!



Demonstrates the use of the BackBeat GreaseMonkey script to uncover an Album object in a catalogue page, then launch a search against the Rdio / Spotify APIs to find the corresponding album in their catalogue.

You could do this for any given Web page, of course, using classic screen scraping, but the use of a standard vocabulary makes it work with no additional effort.



Et voila – reasonable results. If the user has an Rdio subscription, they can immediately start listening and never have to visit your library!

That's Ranganathan's “save the time of the reader” principle right there.

Beyond all of the problems inherent in user-side JavaScript (breaking some sites, introduction of potential memory / CPU issues, breakage when the browser upgrades, etc), it's pretty much impossible to convince most people to install an extension like this.

But it's an extremely interesting proof of concept.

## schema.org: beyond Book

- Object types (sampling):
  - Book, Map, Movie, MusicAlbum, Painting, Photograph, Sculpture
- Attributes (sampling):
  - Main/added entries: author, accountablePerson, contributor, creator, editor
  - Publication: copyrightHolder, copyrightYear, dateCreated, datePublished, publisher
  - Description: description, about, audience, award, contentRating, genre, keywords
  - Social: comment, interactionCount, review

Schema.org defines more than just a few objects and properties (the latter defined at the fairly abstract CreativeWork level) that are of likely interest to libraries.

As of the 1.0a revision, there are 577 different objects defined by the schema.org vocabulary.



## Libraries are objects too

- schema.org vocabulary defines “**Thing > Organization > LocalBusiness > Library**”
- Supports attributes such as:
  - Address and contact info (phone, fax, email)
  - Opening hours
  - Branch relationships
  - Payment types accepted
  - Events
- Search engines could do a lot with this!

Doesn't that make you want to just run out and overhaul your library home page to embed structured data?

## Linking open data

- Authorized LoC headings could link to [http://id.loc.gov/authorities/...](http://id.loc.gov/authorities/)
  - VIAF: interesting target, but does not link out
- What is the canonical link for a book?
  - **isbn** is a property in schema.org, but:
    - The “same” book can have multiple ISBNs
    - Not everything has an ISBN
  - Just link to Freebase? OpenLibrary? WorldCat?
  - FRBR? WEMI?
  - ARGH!?

But Evergreen's current authorities support does not lend itself to validating against or linking to external sources of authorities. *Very interested in what the authorities working group comes up with in the list of desired enhancements!*

Also fits well with Tara Robertson's comments in the Envisioning Evergreen session about teaching the ILS to link outwards instead of inwards.

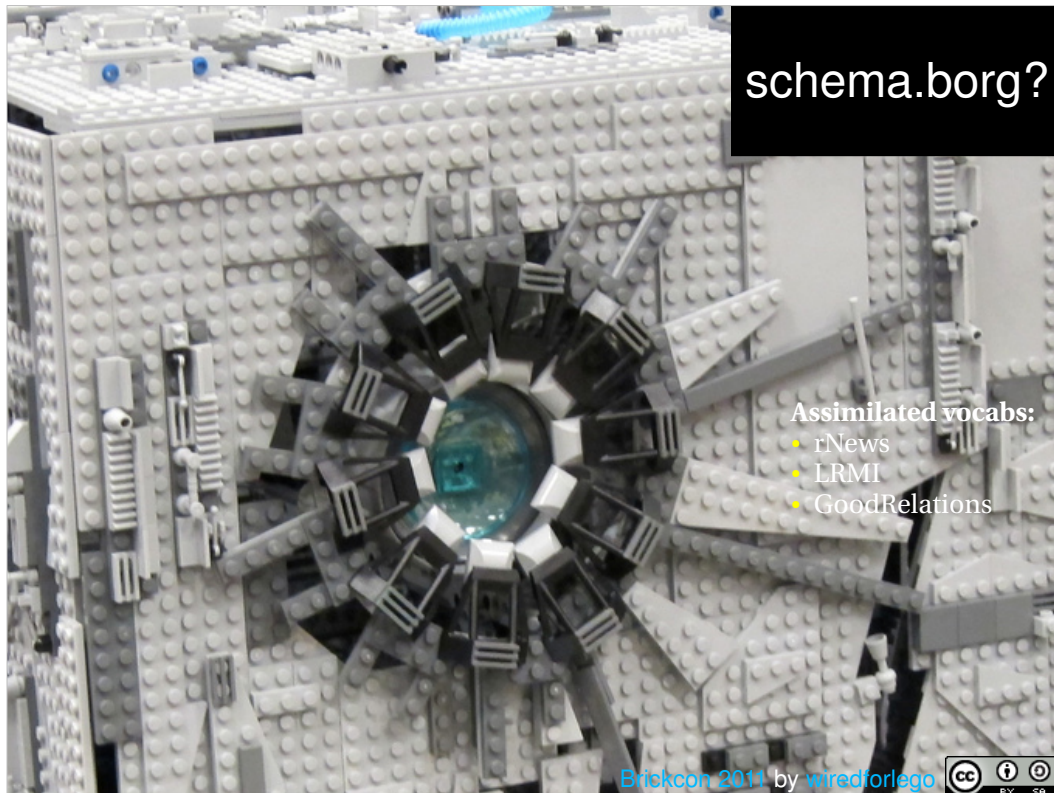
By reaching outwards, we can help search engines reach inwards to us.

## schema.org: vocab in progress

- Extension proposals: coordinated by the W3C [Semantic Web Interest Group](#) Web Schemas task force
- Bibliographic extension proposals:  
[W3C Schema Bib Extend Community Group](#)
  - Initiated by Richard Wallis (OCLC) in September 2012 with monthly calls
  - First action agreed upon March 2013: “to promote citation property to CreativeWork”
  - Much discussion about identifiers, FRBR, BibFrame, and holdings
  - Participants include Richard Wallis, Karen Coyle, Jeff Young, Ross Singer, Jodi Schneider, Laura Dawson...

My participation has (sadly) been very limited since January 2013 due to work commitments. My main interest is in implementation best practices, as that is what I believe the library community would benefit the most from.

To that end, I've been using Evergreen as a base for modelling implementations of schema.org in microdata and RDFa Lite, while OCLC has been tackling the same mission with WorldCat.



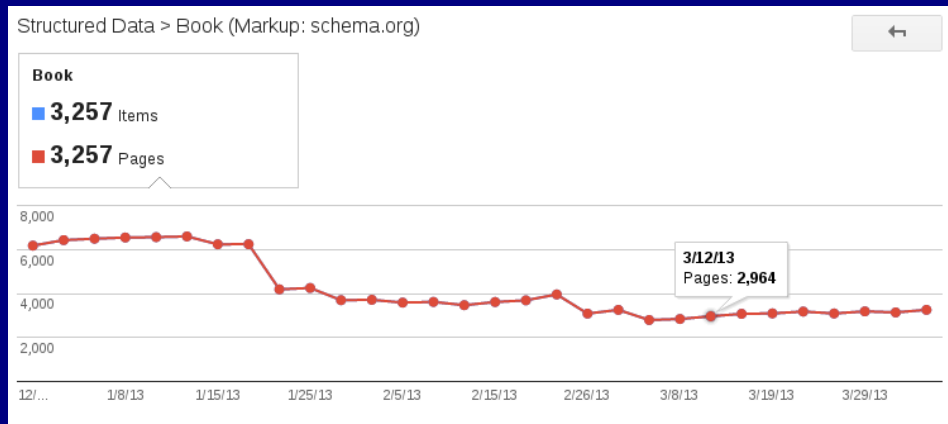
Extensions to schema.org have tended to assimilate existing vocabularies:

Having one place to find the complete ontology that the search engines have blessed is arguably useful.

But this is a bit contrary to the vision of the semantic web that Tim Berners-Lee put forth in 2001 in which different vocabularies would co-exist and evolve independently.

“””Like the Internet, the Semantic Web will be as decentralized as possible. Such Web-like systems generate a lot of excitement at every level, from major corporation to individual user, and provide benefits that are hard or impossible to predict in advance.”””

## A sad tale



I published a site map in late September of all of the public Laurentian University URLs in our Evergreen system.

Google began to crawl it, and (somewhat surprisingly) the number of structured data objects discovered was far below the number of pages crawled.

After approximately 475,000 URLs, Google gave up on crawling any more of our catalogue. The # of structured data objects continued to climb slowly.

In January 2013, the number of objects dropped precipitously. Largely the fault of the MusicRecording vs. Album misclassification, methinks.

Search results would show Laurentian catalogue hits in the top ten for some of the schema.org-recognized content but it is unclear whether this is due to the

## Evergreen: schema.org state

- Evergreen 2.2 through 2.4:
  - Two primary types: Book and MusicRecording\*
  - Just plain text for attributes
- A [working branch](#) improves this greatly:
  - Primary types: Book, Map, MusicAlbum
  - More granular Organization and Person types for main/added entries
  - Birth and death dates for Person objects

To a certain subset of Slashdot users, it is extremely satisfying to cry out “first”, and Evergreen was the first catalogue or discovery layer that I am aware of to release with support for schema.org microdata out of the box. 2.2.0 release in May 2012.

Of interest: OCLC made the same MusicRecording vs. MusicAlbum mistake in their implementation of schema.org in WorldCat, which launched in June 2012.

OCLC treats the entire string “Scott, Dan, 1972-” as a name, whereas Evergreen's working branch breaks that into a name and a birth date

## Evergreen challenges

- Linking out:
  - Works / expressions: where and how?
  - Controlled headings -> linked authorities
- Type mapping from MARC is tough:
  - Is a 2D projected medium a movie, TV show, slide show, or a magic lantern?
- Logic is implemented in TPAC templates
  - That's where we publish HTML, so it makes some sense
  - Also complicates the templates and further hard-codes mappings
- Is it worth the effort?

Linking to Freebase, which covers everything and gives everything an addressable ID (or will generate one, given reasonable metadata) is tempting

Controlled headings in Evergreen currently link to internal IDs for authority records. One might be able to look up the internal authority and resolve that to an external authority, but that seems sub-optimal for performance of a given page. We can probably do better (perhaps use the controlled heading \$0 as a primary key in authority.record\_entry?)

Type mapping probably needs to be handled at a layer above TT2; ideally would be a common effort with other projects.

Immediate payoff is unlikely, but long term... who knows?

## RDFa Lite

- Like Microdata, but a W3C standard that is backwards compatible with RDFa
- Currently only lacks equivalent of `itemref` “copy and paste” attribute
- Supported by schema.org founders, *and* Facebook
- A [working branch](#) offers TPAC templates that are 1:1 with current schema.org microdata

“””

At this point, you may be asking yourself why the two languages are so similar. There is almost 8 years of history here, but to summarize: RDFa was created around the 2004 time frame, Microdata came much later and used RDFa as a design template.

Microdata chose a subset of the original RDFa design to support, but did so in an incompatible way. RDFa Lite then highlighted the subset of the functionality that Microdata did, but in a way that is backwards compatible with RDFa. RDFa Lite did this while keeping the flexibility of the original RDFa intact.

“”” - <http://manu.sporny.org/2012/mythical-differences/>



## HTML + schema.org (RDFa Lite)

```
<div vocab="http://schema.org" typeof="http://schema.org/Book">
  <abbr class="unapi-id" title="tag:localhost,2013:biblio-
record_entry/211" />

  <div><div>
    <h1 property="name">
      Semantic web : ontology and knowledge base enabled tools,
      services, and applications / Amit Sheth, editor.</h1>
    <div>
      <span>
        <a href="/eg/opac/results?query=%20Sheth%20A"
          typeof="http://schema.org/Person"
          property="contributor">
          <span property="name">Sheth, A. (Amit),</span>
          <span property="birthdate">1959</span>
          - </a> (Editor).
        </span>
      </div></div></div></div>
```

Trivial to implement an RDFa Lite variation if you already have microdata in place.

Can use all of schema.org vocabulary; easy to mix in others (FOAF and DC have built-in prefixes, for example).

Perhaps we should cut over to an RDFa Lite implementation?

# References

- Articles:
  - Berners-Lee, T., Hendler, J., and Lassila, O. (2001) "The Semantic Web". Scientific American, [May](#).
  - Coyle, Karen. "Rich Snippets". Coyle's InFormation. <http://kcoyle.blogspot.ca/2012/09/rich-snippets.html>
  - Sporny, Manu. "Mythical Differences: RDFa Lite vs. Microdata." The Beautiful, Tormented Machine. <http://manu.sporny.org/2012/mythical-differences/>
- Sites:
  - schema.org ontology: <http://schema.org>
  - RDFa Lite standard: <http://www.w3.org/TR/rdfa-lite/>
  - W3C schema.org bibliographic extension community: <http://www.w3.org/community/schemabibex>
  - W3C semantic web community: <http://www.w3.org/standards/semanticweb>
- Tools:
  - BackBeat: <https://github.com/rsinger/BackBeat>

Note: now that I've steeped myself in structured data, it feels ridiculous to put together a traditional presentation and focus primarily on presentation. Ah well... Next time, HTML5 to the rescue.